Electronic Journal of Statistics
Vol. 16 (2022) 3672–3694
ISSN: 1935-7524
https://doi.org/10.1214/21-EJS1976

# The empirical saddlepoint estimator

**Benjamin Holcblat**

*University of Luxembourg, FDEF,*
*6 Rue Richard Coudenhove-Kalergi, L-1359, Luxembourg e-mail:* Benjamin.Holcblat@uni.lu

**Fallaw Sowell**

*Carnegie Mellon University,*
*Tepper School of Business,*
*5000 Forbes Ave. Pittsburgh, PA, USA e-mail:* fs0v@andrew.cmu.edu

**Abstract:** We define a moment-based estimator that maximizes the empirical saddlepoint (ESP) approximation of the distribution of solutions to empirical moment conditions. We call it the ESP estimator. We prove its existence, consistency and asymptotic normality, and we propose novel test statistics. We also show that the ESP estimator corresponds to the MM (method of moments) estimator shrunk toward parameter values with lower implied estimated variance, so it reduces the documented instability of existing moment-based estimators. In the case of just-identified moment conditions, which is the case we focus on, the ESP estimator is different from the MM estimator, unlike the more recent alternatives, such as the empirical-likelihood-type estimators.

**Keywords and phrases:** Empirical saddlepoint approximation, method of moments, weak instruments, Kullback-Leibler divergence criterion, variance penalization.

Received October 2020.

## Contents

## 1. Introduction

The saddlepoint (SP) approximation has been developed to approximate distributions. Because of its accuracy, it is regularly used in several fields, such as numerical analysis (e.g., [57]'s algorithm to approximate binomial distributions, and which is notably used in the statistical software R) and actuarial sciences (e.g., [21]'s approximation for distributions tails). In statistics and econometrics, the SP approximation and its empirical version —the empirical saddlepoint (ESP) approximation— have been used to approximate finite-sample distributions [e.g., 15, 42, 16]. Standard monographs and introductions about the ESP and the SP approximation for statistics include [28], [52], [47], [30] and [8].

In the present paper, we propose to use the ESP approximation to define a *point* estimator $\hat{\theta}_T$. We call it the ESP estimator, and denote it with $\hat{\theta}_T$. It maximizes the [72]'s ESP approximation of the distribution of solutions to the empirical moment conditions

$$\frac{1}{T}\sum_{t=1}^{T}\psi\left(X_t,\theta\right)=0 \tag{1}$$

where $\psi(.,.)$ denotes the moment function s.t. $\mathbb{E}[\psi(X_1,\theta_0)] = 0_{m\times 1}$ an $m$-dimensional vector of zeros, $(X_t)_{t=1}^{T}$ i.i.d. data, $\theta_0 \in \boldsymbol{\Theta} \subset \mathbf{R}^m$ the unknown parameter of interest, and $T$ the sample size. In the present paper, for clarity and simplicity, in line with most of the SP literature, we consider the so-called just-identified case, in which the number of parameters is the same as the number of moment conditions.

The ESP estimator is a moment-based estimator. Since [63, 64]'s method of moment (MM), moment-based estimators have been found useful in a variety of applications (e.g., covariance structure analysis in psychology, and asset pricing in economics). Their two main advantages are (i) they do not require a parametric family of probability distributions for the data so they are less prone to model misspecification, and (ii) they allow complex models for which the likelihood function is intractable.

Nevertheless, the increase use of the MM and its extensions has revealed that they can be unstable and perform poorly in finite samples (e.g., July 1996 special issue of Journal of Business & Economics Statistics). The idea of the ESP estimator $\hat{\theta}_T$ is to improve on the MM estimator $\theta_T^*$ as follows — note the difference between $\hat{\theta}_T$, which denotes the ESP estimator, and $\theta_T^*$, which denotes the MM estimator or, equivalently, a solution to the empirical moment conditions (1). By definition, the MM estimate $\theta_T^*(\omega)$ solves the empirical moment conditions (1) evaluated for a given sample $(X_t(\omega))_{t=1}^{T}$, but it typically does not solve the empirical moment condition for another sample $(X_t(\dot{\omega}))_{t=1}^T$ where $\dot{\omega} \neq \omega$. Thus, we might want an estimate that not only takes into account the empirical moment conditions evaluated for the given sample $(X_t(\omega))_{t=1}^T$, but also for other potential samples. More precisely, we want an estimate that accounts for all possible evaluations of the empirical moment conditions according to their probability of occurrence. This is what the ESP estimate does: The ESP estimate maximizes the ESP approximation to the distribution of the solutions to the empirical moment conditions [72]. If the empirical moment conditions (1) have a unique solution with a continuous distribution, the ESP estimator maximizes the ESP approximation of a probability density function of the solution $\theta_T^*$. We rely on the ESP approximation because simulation and theoretical evidence shows the ESP approximation can be very accurate in small sample [e.g., 16, 72].

We also show that the ESP estimator corresponds to an MM estimator shrunk toward parameter values with lower implied estimated variance of the solution to the corresponding finite-sample moment conditions. More precisely, we decompose the logarithm of the ESP approximation as the sum of a term, which is maximized at the MM estimator, and a variance penalty, which discounts any parameter value $\dot{\theta} \in \mathbf{\Theta}$ that implies a high estimated variance for the corresponding MM estimator. Under assumptions adapted from the entropy literature, we establish the ESP estimator has the same good asymptotic properties as the MM estimator, so the variance penalization is a finite-sample correction. We also derive the ESP counterparts of the Wald, Lagrange multiplier (LM), analogue likelihood-ratio (ALR) test statistics, as well as another test statistic. Then, we investigate the ESP estimator through Monte-Carlo simulations. We compare its performance with the exponential tilting (ET) estimator, which is equal to the MM estimator in the just-identified case. Results show that the variance penalization of the ESP estimator reduces the finite-sample instability of the ET estimator (or equivalently, of the MM estimator). An empirical application illustrates the gain from this greater stability in terms of inference.

The ESP estimator is not the first proposal to improve on the MM and its extensions. Alternative moment-based approaches have been proposed such as the empirical likelihood approach of Owen [67], the continuously updating approach [37], the already-mentioned exponential tilting (ET) approach [51, 44], and combinations of the aforementioned approaches [e.g., 73]. All these approaches yield an estimator closely related to the empirical likelihood estimator, so we call them empirical-likelihood-type estimators. In the just-identified case, when well-defined, all of these empirical-likelihood-type estimators are numerically equal to the original Pearson's MM estimator $\theta_T^*$. Because we focus on the just-identified case, it is sufficient for us to compare the ESP estimator with the MM estimator, or with one of any of these more recent estimators.

In addition to the already cited papers, the present paper, which supersedes the unpublished manuscript [78], is related to many other ones. We clarify these relations in Section 5 (p. 3687). To the best of our knowledge, none of the prior papers use the SP or the ESP to propose a novel moment-based point estimator. Overall, the present paper (i) brings together the literature on the saddlepoint approximation and the literature on moment-based estimation, and (ii) points out the untapped potential of the ESP to tackle issues faced in moment-based estimation.

**Remark 1.** Another motivation for the ESP estimator is decision theoretic. The ESP estimator follows from the minimization of the expectation of a loss "function" that equals zero when $\theta$ solves the empirical moment conditions and one otherwise by normalization. This motivation is similar to the decision-theoretic justification for the Bayesian maximum a posteriori estimator [69]. As in Bayesian analysis, the choice of other loss functions is possible. The investigation of different loss functions is left for future research.

## 2. Finite-sample analysis

In the present section, we remind the formula for the ESP approximation, and analyze its finite-sample structure. Then, we decompose the log-ESP into two terms and show that the ESP estimator is a MM estimator shrunk toward parameter values with lower implied estimated variance, so the estimation stability is improved.

### 2.1. The ESP approximation

Formalizing and generalizing prior works [16, 26, 85, 86], [72] propose the following ESP approximation to estimate the distribution of a solution to the empirical moment conditions (1)

$$\hat{f}_{\theta_T^*}(\theta) := \exp\left\{T \ln\left[\frac{1}{T}\sum_{t=1}^{T} e^{\tau_T(\theta)'\psi_t(\theta)}\right]\right\}\left(\frac{T}{2\pi}\right)^{m/2} |\Sigma_T(\theta)|_{\det}^{-\frac{1}{2}} \tag{2}$$

where $|.|_{\det}$ denotes the determinant function, $\theta_T^*$ a solution to (1), $\psi_t(.) := \psi(X_t, .)$, and

$$\begin{aligned}
\Sigma_T(\theta) \quad &:= \quad \left[\sum_{t=1}^{T} w_{t,\theta}\frac{\partial \psi_t(\theta)}{\partial \theta'}\right]^{-1}\left[\sum_{t=1}^{T} w_{t,\theta}\psi_t(\theta)\psi_t(\theta)'\right] \\
&\qquad \times \left[\sum_{t=1}^{T} w_{t,\theta}\frac{\partial \psi_t(\theta)'}{\partial \theta}\right]^{-1}, && (3) \\
w_{t,\theta} \quad &:= \quad \frac{\exp\left[\tau_T(\theta)'\psi_t(\theta)\right]}{\sum_{i=1}^{T}\exp\left[\tau_T(\theta)'\psi_i(\theta)\right]}, && (4) \\
\tau_T(\theta) \quad &\text{such that} \quad \sum_{t=1}^{T}\psi_t(\theta)\frac{\exp\left[\tau_T(\theta)'\psi_t(\theta)\right]}{\sum_{i=1}^{T}\exp\left[\tau_T(\theta)'\psi_i(\theta)\right]} \times \frac{1}{T} = 0. && (5)
\end{aligned}$$

The ESP estimator maximizes the ESP approximation (2), or equivalently, its logarithm, which is given in the upcoming formula (7) apart for terms constant w.r.t. (with respect to) $\theta$. The ESP approximation (2) is the empirical counterpart of the SP approximation of [27]. From a computational point of view, the ESP approximation (2) is not complicated. See Section 4.1 and the Appendix section 10 in [41] for more details. The only implicit quantity is $\tau_T(\theta)$, which solves the tilting equation (5), which, in turn, is just the FOC (first-order condition) of the unconstrained convex problem $\min_{\tau\in\mathbf{R}^m}\sum_{t=1}^T e^{\tau'\psi_t(\theta)}$. A full understanding of the ESP approximation (2) arguably requires to work through higher-order asymptotic expansions along the lines of [27]. However, direct inspection of the ESP approximation (2) also provides insight for how it incorporates information from the data through two channels.

The first channel is the *ET (exponential tilting) term* $\exp\left\{T\ln\left[\frac{1}{T}\sum_{t=1}^T e^{\tau_T(\theta)'\psi_t(\theta)}\right]\right\}$. In equation (5), for any $\theta\in\mathbf{\Theta}$, the terms $\frac{\exp\left[\tau_T(\theta)'\psi_t(\theta)\right]}{\sum_{i=1}^T\exp[\tau_T(\theta)'\psi_i(\theta)]}$ tilt the empirical probability $1/T$, so the finite-sample moment conditions (5) hold. This tilting determines, through equation (4), the multinomial distribution $(w_{t,\theta})_{t=1}^T$ that is the closest to the empirical distribution —in the sense of the Kullback-Leibler divergence criterion— s.t. the finite-sample moment conditions (5) holds: The tilting equation (5) is the FOC w.r.t. (with respect to) $\tau$ of the Lagrangian dual problem of the minimization problem

$$\min_{(w_{1,\theta},w_{2,\theta},\cdots,w_{T,\theta})\in]0,1]^T}\sum_{t=1}^T w_{t,\theta}\log\left(\frac{w_{t,\theta}}{1/T}\right)$$
$$\text{s.t.}\sum_{t=1}^T w_{t,\theta}\psi_t(\theta)=0\text{ and }\sum_{t=1}^T w_{t,\theta}=1, \tag{6}$$

where $\sum_{t=1}^T w_{t,\theta}\log[w_{t,\theta}/(1/T)]$ is the Kullback-Leibler divergence criterion between the empirical distribution and the multinomial distribution $(w_{t,\theta})_{t=1}^T$ with the same support [e.g., 13, 19]. Then, for the given $\theta\in\mathbf{\Theta}$, in the ESP approximation 2, the ET term $\exp\left\{T\ln\left[\frac{1}{T}\sum_{t=1}^T e^{\tau_T(\theta)'\psi_t(\theta)}\right]\right\}$ indicates the extent of the tilting needed to set the finite-sample moment conditions (6) (or equivalently, equation (5)) to zero. The bigger is the tilting of the empirical distribution, the less compatible are the data with $\theta$ solving the empirical moment conditions, and the smaller should be the ET term $\exp\left\{T\ln\left[\frac{1}{T}\sum_{t=1}^T e^{\tau_T(\theta)'\psi_t(\theta)}\right]\right\}$. It can be easily seen that $\frac{1}{T}\sum_{t=1}^T e^{\tau_T(\theta)'\psi_t(\theta)}$ reaches its maximum when $\theta$ is a solution $\theta_T^*$ of the empirical moment conditions (1), i.e., when $\tau_T(\theta_T^*)=0_{m\times 1}$ and no tilting is needed —For a formal proof, one can follow the same reasoning as in the proof of Lemma 10 (Appendix) with the empirical distribution in lieu of $\mathbb{P}$.

In the ESP approximation on equation (2), the second term $\left(\frac{T}{2\pi}\right)^{m/2}$ comes from the multivariate Gaussian distribution that is the leading term of the Edgeworth's asymptotic expansions underlying ESP approximations. However, because it is constant w.r.t. $\theta$, it does not affect the maximization of the ESP approximation, so it is *not* an information channel for the ESP estimator.

The remaining term $|\Sigma_T(\theta)|_{\det}^{-\frac{1}{2}}$, which we call the *variance term*, is the second channel through which the ESP approximation incorporates information from data. The variance term discounts the ET term according to the tilted estimated variance of the solution to the finite-sample moment conditions. Under standard assumptions, a consistent estimator of the asymptotic variance of $\sqrt{T}(\theta_T^* - \theta_0)$ is $\Sigma_T(\theta_T^*) := \left[\sum_{t=1}^T w_{t,\theta_T^*} \frac{\partial \psi_t(\theta_T^*)}{\partial \theta}\right]^{-1} \left[\sum_{t=1}^T w_{t,\theta_T^*} \psi_t(\theta_T^*) \psi_t(\theta_T^*)'\right] \times \left[\sum_{t=1}^T w_{t,\theta_T^*} \frac{\partial \psi_t(\theta_T^*)'}{\partial \theta}\right]^{-1}$. The bigger the variance term is, the less plausible a solution takes exactly this value, and the smaller is $|\Sigma_T(\theta)|_{\det}^{-\frac{1}{2}}$ —note the negative power. Therefore, overall, for a given $\theta \in \mathbf{\Theta}$, the bigger the tilting or the estimated variance, the smaller the ESP approximation, i.e., the estimated probability weight $\hat{f}_{\theta_T^*}(\theta)$ that $\theta$ solves the empirical moment conditions (1).

## *2.2. The ESP estimator as a shrinkage estimator*

As explained in the introduction, the more recent moment-based estimators are numerically equal to the Pearson's MM estimator in the just-identified case. Thus, it is sufficient to compare the ESP estimator with one of them in order to understand the difference between the former and the other proposed moment-based estimators. The ET estimator of [51] and [44] is particularly convenient for this purpose. Taking the logarithm of the ESP approximation (2), removing the terms constant w.r.t. $\theta$, and dividing by the sample size $T$, it can be seen that the ESP estimator $\hat{\theta}_T$ maximizes the objective function

$$\ln \left[\frac{1}{T} \sum_{t=1}^{T} e^{\tau_T(\theta)' \psi_t(\theta)}\right] - \frac{1}{2T} \ln |\Sigma_T(\theta)|_{\det}, \tag{7}$$

where $\ln \left[\frac{1}{T} \sum_{t=1}^T e^{\tau_T(\theta)' \psi_t(\theta)}\right]$ is a strictly increasing transformation of the objective function of the ET estimator. Thus, the difference between the ESP estimator and the ET estimators comes only from the log-variance term $-\frac{1}{2T}$ $\ln |\Sigma_T(\theta)|_{\det}$. As explained in Section 2.1, the variance term incorporates additional information, which penalizes parameter values with higher implied estimated variance. More precisely, the variance term discounts any parameter value $\dot{\theta} \in \mathbf{\Theta}$ that implies a high estimated variance for the ET —or equivalently, MM— estimator of $\dot{\theta}$ based on the tilted moment condition $\int_{\mathbf{\Omega}} \psi(X_1(\omega), \dot{\theta}) \mathrm{P}_{\dot{\theta}}$ $(\mathrm{d}\omega) = 0_{m \times 1}$, where $\frac{\mathrm{dP}_{\dot{\theta}}}{\mathrm{d}\mathbb{P}} := \frac{e^{\tau(\dot{\theta})' \psi(X_1, \dot{\theta})}}{\mathbb{E}[e^{\tau(\dot{\theta})' \psi(X_1, \dot{\theta})}]}$ with $\mathbb{P}$ the physical probability measure. Thus, the ESP estimator is an ET estimator shrunk toward parameter values with lower implied estimated variance. As the following Proposition 1 shows, it immediately implies a smaller estimated variance for the ESP estimator.

**Proposition 1** (Shrunk estimated variance of the ESP estimator)**.** *Assume the existence of an ESP estimator* $\hat{\theta}_T$ *and an ET estimator* $\theta_T^*$ *s.t.* $\hat{\theta}_T$ *is different* *from any ET estimator i.e.,* $\hat{\theta}_T \notin \arg\max_{\theta\in\mathbf{\Theta}} \ln\left[\frac{1}{T}\sum_{t=1}^{T} e^{\tau_T(\theta)'\psi_t(\theta)}\right]$. *Then*

$$|\Sigma_T(\hat{\theta}_T)|_{\det} < |\Sigma_T(\theta_T^*)|_{\det}.$$

For more details about Proposition 1, see the Appendix section 9. The variance shrinkage is desirable because of the documented instability of existing nonlinear moment-based estimators.

## 3. Asymptotic properties

In the present section, we investigate the asymptotic properties of the ESP estimator. Good asymptotic properties can be regarded as a minimal requirement for the ESP estimator, which is based on a small-sample asymptotic approximation. While the asymptotic properties of the ESP estimator are standard, their complete proofs, which are in the Appendix, require the development of some nonstandard arguments.

### 3.1. Existence, consistency and asymptotic normality

We require the following assumption to prove the existence and the consistency of the ESP estimator.

**Assumption 1. (a)** *The data* $(X_t)_{t=1}^{\infty}$ *are a sequence of i.i.d. random vectors of dimension p on the complete probability sample space* $(\mathbf{\Omega}, \mathcal{E}, \mathbb{P})$. **(b)** *Let the moment function* $\psi : \mathbf{R}^p \times \mathbf{\Theta}^\epsilon \mapsto \mathbf{R}^m$ *be s.t.* $\theta \mapsto \psi(X_1, \theta)$ *is continuously differentiable* $\mathbb{P}$*-a.s., and* $\forall\theta \in \mathbf{\Theta}^\epsilon$, $x \mapsto \psi(x, \theta)$ *is* $\mathcal{B}(\mathbf{R}^p)/\mathcal{B}(\mathbf{R}^m)$*-measurable, where, for* $\epsilon > 0$, $\mathbf{\Theta}^\epsilon$ *denotes the* $\epsilon$*-neighborhood of* $\mathbf{\Theta}$*, and* $\mathcal{B}(\mathbf{R}^p)$ *the Borel* $\sigma$*-algebra on* $\mathbf{R}^p$. **(c)** *In the parameter space* $\mathbf{\Theta}$*, there exists a unique* $\theta_0 \in \operatorname{int}(\mathbf{\Theta})$ *s.t.* $\mathbb{E}[\psi(X_1, \theta_0)] = 0_{m\times 1}$. **(d)** *Let the parameter space* $\mathbf{\Theta} \subset \mathbf{R}^m$ *be a compact set, s.t., for all* $\theta \in \mathbf{\Theta}$*, there exists* $\tau(\theta) \in \mathbf{R}^m$ *that solves the equation* $\mathbb{E}\left[e^{\tau'\psi(X_1,\theta)}\psi(X_1, \theta)\right] = 0$ *for* $\tau$. **(e)** $\mathbb{E}\left[\sup_{(\theta,\tau)\in\mathbf{S}^\epsilon} e^{2\tau'\psi(X_1,\theta)}\right] < \infty$ *where* $\mathbf{S} := \{(\theta, \tau) : \theta \in \mathbf{\Theta}\ \&\ \tau \in \mathbf{T}(\theta)\}$ *and* $\mathbf{T}(\theta) := \overline{B_{\epsilon_\mathbf{T}}(\tau(\theta))}$ *with* $\overline{B_{\epsilon_\mathbf{T}}(\tau(\theta))}$ *the closed ball of radius* $\epsilon_\mathbf{T} > 0$ *and center* $\tau(\theta)$. **(f)** $\mathbb{E}\left[\sup_{\theta\in\mathbf{\Theta}} |\frac{\partial\psi(X_1,\theta)}{\partial\theta'}|^2\right] < \infty$, *where* $|.|$ *denotes the Euclidean norm.* **(g)** $\mathbb{E}\left[\sup_{\theta\in\mathbf{\Theta}^\epsilon} |\psi(X_1, \theta)\psi(X_1, \theta)'|^2\right] < \infty$. **(h)** *For all* $\theta \in \mathbf{\Theta}$, $\Sigma(\theta) := \left[\mathbb{E}e^{\tau(\theta)'\psi(X_1,\theta)}\frac{\partial\psi(X_1,\theta)}{\partial\theta'}\right]^{-1} \mathbb{E}\left[e^{\tau(\theta)'\psi(X_1,\theta)}\psi(X_1, \theta)\psi(X_1, \theta)'\right] \left[\mathbb{E}e^{\tau(\theta)'\psi(X_1,\theta)}\frac{\partial\psi(X_1,\theta)'}{\partial\theta}\right]^{-1}$ *is invertible.*

We require the following additional assumption to prove the asymptotic normality of the ESP estimator.

**Assumption 2. (a)** *The function* $\theta \mapsto \psi(X_1, \theta)$ *is three times continuously differentiable in a neighborhood* $\mathcal{N}$ *of* $\theta_0$ *in* $\mathbf{\Theta}$ $\mathbb{P}$*-a.s.* **(b)** *There exists a* $\mathcal{B}(\mathbf{R}^p)/\mathcal{B}(\mathbf{R})$*-measurable function* $b(.)$ *satisfying* $\mathbb{E}\left[\sup_{\theta\in\mathcal{N}}\sup_{\tau\in\mathbf{T}(\theta)} e^{k_1\tau'\psi(X_1,\theta)}b(X_1)^{k_2}\right] <$

*$\infty$ for $k_1 \in [\![1,2]\!]$ and $k_2 \in [\![1,4]\!]$ s.t., for all $j \in [\![0,3]\!]$, $\sup_{\theta\in\mathcal{N}} |\nabla^j \psi(X_1,\theta)| \leqslant b(X_1)$ where $\nabla^j \psi(X_1,\theta)$ denotes a vector of all partial derivatives of $\theta \mapsto \psi(X_1,\theta)$ of order $j$, and $[\![a,b]\!] := [a,b] \cap \mathbf{Z}$ for all $(a,b) \in \mathbf{R}^2$.*

Assumptions 1 and 2 are stronger than the usual assumptions in the MM literature, but are similar to assumptions used in the entropy literature and related literatures. Assumptions 1 and 2 are essentially adapted from [33, 51], and [73]. See also [12] for similar assumptions. Section 7.1 of the Appendix contains a detailed discussion of Assumptions 1 and 2. Under Assumptions 1 and 2, the following theorem establishes the existence, the strong consistency, and the asymptotic normality of the ESP estimator $\hat{\theta}_T$.

**Theorem 1** (Existence, consistency and asymptotic normality)**.** *Under Assumption 1, $\mathbb{P}$-a.s. for $T$ big enough, there exists $\hat{\theta}_T$ s.t.*

*(i) $\mathbb{P}$-a.s. as $T \to \infty$, $\hat{\theta}_T \to \theta_0$; and*

*(ii) under the additional Assumption 2, as $T \to \infty$, $\sqrt{T}(\hat{\theta}_T - \theta_0) \xrightarrow{D} \mathcal{N}(0, \Sigma(\theta_0))$.*

*where $\Sigma(\theta_0) := \left[\mathbb{E}\frac{\partial\psi(X_1,\theta_0)}{\partial\theta'}\right]^{-1} \mathbb{E}\left[\psi(X_1,\theta_0)\psi(X_1,\theta_0)'\right] \left[\mathbb{E}\frac{\partial\psi(X_1,\theta_0)'}{\partial\theta}\right]^{-1}$, $\xrightarrow{D}$ denotes the convergence in distribution.*

Theorem 1 shows the variance penalization $-\frac{1}{2T}\ln|\Sigma_T(\theta)|_{\text{det}}$ vanishes sufficiently quickly asymptotically, so it does not distort the first-order asymptotic of the estimator. In particular, Theorem 1(ii) shows that the ESP estimator reaches the same semiparametric efficiency bound as the MM and the more recent moment-based estimators [11, 10]. Parts of the proofs of Theorem 1 are involved, although the proofs strategies follow traditional approaches. The proof of existence follows the Schmetterer and Jennrich approach ([74] Chap. 5; [45]), with an additional complication coming from the implicit nature of the function $\theta \mapsto \tau_T(\theta)$. The proof of Theorem 1(i) (i.e., consistency) follows Wald's approach to consistency [84]. The usual consistency approach for empirical-likelihood-type estimators [62, 77] cannot be easily followed here because of the variance term. Moreover, the later consistency approach does not articulate well with an existence proof. The basic idea of our consistency proof is to show that, $\mathbb{P}$-a.s. for $T$ big enough, the ESP estimator maximizes the LogESP function (7), where, $\mathbb{P}$-a.s. as $T \to \infty$,

$$\sup_{\theta\in\mathbf{\Theta}} \left| \ln\left[\frac{1}{T}\sum_{t=1}^{T} e^{\tau_T(\theta)'\psi_t(\theta)}\right] - \ln\mathbb{E}[e^{\tau(\theta)'\psi(X_1,\theta)}] \right| = o(1) \text{ , and}$$
$$\sup_{\theta\in\mathbf{\Theta}} \left| \frac{1}{2T}\ln|\Sigma_T(\theta)|_{\text{det}} \right| = O(T^{-1}). \tag{8}$$

The two main complications w.r.t. the proofs available in the entropy literature are the following. First, the need to ensure that, for $T$ big enough, for all $\theta \in \mathbf{\Theta}$, $|\Sigma_T(\theta)|_{\text{det}}$ is bounded away from zero, so that the LogESP function (7) does not diverge on parts of the parameter space. Second, we establish that the joint parameter space for $\theta$ and $\tau$ (i.e., $\mathbf{S}$) is a compact set. For the latter purpose, we appropriately modify assumptions from the entropy literature and develop a proof based on set-valued analysis. The proof of Theorem 1(ii) (i.e., asymptotic normality) follows the traditional approach of expanding the FOCs. The two main complications w.r.t. the proofs in the entropy literature are the following. First, instead of expanding the exact FOC $\frac{\partial \ln[\hat{f}_{\theta_T^*}(\theta)]}{\partial\theta}\Big|_{\theta=\hat{\theta}_T}$, we expand an approximate FOC combined with the FOC (5) for $\tau$. This is technical because it involves differentiation of the log variance term. Second, we control the asymptotic behaviour of the derivatives that come from the log-variance term $\ln|\Sigma_T(\theta)|_{\det}$. Another shorter proof approach based on an approximate FOC of the first term of the log-ESP is possible. We do not follow it because it would complicate and lengthen the presentation and the proof of the upcoming Theorem 2.

### *3.2. More on inference: The trinity+1*

The ESP estimator provides different ways to test parameter restrictions

$$\mathrm{H}_0 : r(\theta_0) = 0_{q\times 1} \tag{9}$$

where $r : \mathbf{\Theta} \to \mathbf{R}^q$ with $q \in [\![1, \infty[\![$. More precisely, within the ESP framework, there exist the usual trinity of Wald, LM and ALR tests statistics, plus another test statistic, which we call the Tilt test statistic.

In addition to Assumptions 1 and 2, we require the following standard and mild assumption to establish the asymptotic distribution of the Wald, LM, ALR, and Tilt statistics.

**Assumption 3** (For the trinity+1)**. (a)** *The function $r : \mathbf{\Theta} \to \mathbf{R}^q$ in the null hypothesis (9) is continuously differentiable.* **(b)** *The derivative $R(\theta) := \frac{\partial r(\theta)}{\partial \theta'}$ is full rank at $\theta_0$.*

Under Assumptions 1, 2 and 3, the following theorem shows that the Wald, LM ALR, and Tilt statistics asymptotically follow a chi-squared distribution with $q$ degrees of freedom.

**Theorem 2** (The trinity+1: Wald, LM, ALR and Tilt tests)**.** *Define $R(\theta) := \frac{\partial r(\theta)}{\partial \theta'}$, and the following Wald, LM, ALR and Tilt test statistics*

$$\begin{aligned}
\mathrm{Wald}_T &:= Tr(\hat{\theta}_T)'[R(\hat{\theta}_T)\widehat{\Sigma(\theta_0)}_T R(\hat{\theta}_T)']^{-1} r(\hat{\theta}_T) \\
\mathrm{LM}_T &:= T\check{\gamma}_T'[R(\check{\theta}_T)\widehat{\Sigma(\theta_0)}_T R(\check{\theta}_T)']\check{\gamma}_T = \frac{1}{T}\frac{\partial \ln[\hat{f}_{\theta_T^*}(\check{\theta}_T)]}{\partial \theta'}\widehat{\Sigma(\theta_0)}_T \frac{\partial \ln[\hat{f}_{\theta_T^*}(\check{\theta}_T)]}{\partial \theta} \\
\mathrm{ALR}_T &:= 2\{\ln[\hat{f}_{\theta_T^*}(\hat{\theta}_T)] - \ln[\hat{f}_{\theta_T^*}(\check{\theta}_T)]\} \\
\mathrm{Tilt}_T &:= T\tau_T(\check{\theta}_T)'\widehat{V}_T \tau_T(\check{\theta}_T)
\end{aligned}$$

*where $\widehat{\Sigma(\theta_0)}_T$ and $\widehat{V}_T$ are symmetric matrices that converge in probability to $\Sigma(\theta_0)$ and $\mathbb{E}[\psi(X_1, \theta_0)\psi(X_1, \theta_0)']$, respectively; and where $\check{\gamma}_T$ and $\check{\theta}_T$ respectively* *denote the Lagrange multiplier and a solution to the maximization of* $\hat{f}_{\theta_T^*}(\theta)$ *w.r.t.* $\theta \in \boldsymbol{\Theta}$ *under the constraint that* $r(\theta) = 0_{q\times 1}$*, i.e.,* $\check{\theta}_T \in \arg\max_{\theta\in\check{\Theta}} \hat{f}_{\theta_T^*}(\theta)$ *with* $\check{\Theta} := \{\theta \in \Theta : r(\theta) = 0_{q\times 1}\}$ *and* $\check{\gamma}_T$ *s.t.* $\frac{1}{T}\frac{\partial \ln[\hat{f}_{\theta_T^*}(\check{\theta}_T)]}{\partial\theta} + \frac{\partial r(\check{\theta}_T)'}{\partial\theta}\check{\gamma}_T = 0_{m\times 1}$. *Under Assumptions 1, 2 and 3, if the test hypothesis* (9) *holds, as* $T \to \infty$*,*

$$\mathrm{Wald}_T, \mathrm{LM}_T, \mathrm{ALR}_T, \mathrm{Tilt}_T \xrightarrow{D} \chi_q^2.$$

Theorem 2 can also be used to obtain valid confidence regions by the inversion of the test statistics with $\check{\theta}_T = \theta_0$. Our Wald, LM, ALR and Tilt test statistics share some similarity with the test statistics proposed in [51], [44], and [70]. As explained in Section 2, the difference between our test statistic and aforementioned test statistics come from the variance term, which affects both the objective function and the (possibly constrained) estimator. Thus, an inspection of the formulas for the test statistics shows the LM and ALR test statistics are the most different from their ET counterpart, and that the ALR test statistic exploits more the variance term than the LM test statistic—a derivative of a function contains less information than the function. Moreover, as usual, LM test statistics should be avoided in non linear setting because of local extrema. The proof of Theorem 2 follows the traditional proof strategy for deriving the trinity. The main complications w.r.t. the proofs available in the entropy literature are the same as for the proof of Theorem 1(ii). Note that the uniform convergences (8) of the two parts of the ESP objective function combined with results from the entropy literature do not imply Theorem 2 because the trinity+1 test statistics are scaled by $T$, and $T\frac{1}{2T}\ln|\Sigma_T(\theta)|_{\det} \neq o(1)$, $\mathbb{P}$-a.s. as $T \to \infty$.

**Remark 2.** As a referee noted, among [51], [44], and [70], the latter is unique to establish a relative error of order $O(T^{-1})$ for the SP approximation. A relative error of lower magnitude is especially useful for improving accuracy in distributions tails, which are of particular interest for testing.

## 4. Examples

In the present section, we further investigate and illustrate the finite-sample properties of the ESP estimator. We focus on the comparison with the ET estimator, as previously noted, (i) in the just-identified case, which is the case addressed in the present paper, the MM estimator and the more recent moment-based estimators are equal to the ET estimator so there is no loss of generality in terms of point estimation, and (ii) the ESP objective function nests the ET objective function, so that the source of the difference between the two is easily understood —it necessarily comes from the variance term (see Section 2.2). In order to provide some finite-sample evidence for the test statistics, we also report their actual rejection probabilities. More information and simulation results are presented in the Appendix section 10.

**Remark 3.** In addition to our finite-sample analysis of the ESP objective function (Section 2), our derivation of the first-order asymptotic properties (Section 3), our Monte-Carlo simulations and empirical application (present section), another way to shed light on the finite-sample properties of the ESP estimator would be to derive its higher-order asymptotic properties such as its second-order bias [e.g., 68]. However, higher-order asymptotic properties of shrinkage estimators are typically complex to study, and have often remained an open problem. The ESP estimator appears to be no exception among shrinkage estimators. Our preliminary derivations yield a long and complicated structure for the second-order bias, from which we struggle to gain insight. The length and the complexity of the second-order bias mainly comes from (i) the derivatives of the variance $|\Sigma_T(\theta)|_{\det}^{-1/2}$; and (ii) the reliance on the exact FOCs instead of approximate FOCs. A mild preview of this complexity can be seen in the proof of asymptotic normality.

### *4.1. Numerical example: Monte-Carlo simulations*

#### *4.1.1. ET and ESP estimators for the two-parameter Hall and Horowitz model*

We simulate a two-parameter just-identified version of the [35] model, which has become a standard benchmark to compare the performance of moment-based estimators in statistics [e.g., 73, 56] and econometrics [e.g., 44, 50]. This model can be interpreted as a simplified consumption-based asset pricing model where $\beta$ is the relative risk aversion (RRA) parameter [31]. In the simulations, we estimate the two parameters $(\mu, \beta)$ with the moment function

$$\psi_t(\beta, \mu) = \begin{bmatrix} \exp\{\mu - \beta(X_t + Y_t) + 3Y_t\} - 1 \\ Y_t(\exp\{\mu - \beta(X_t + Y_t) + 3Y_t\} - 1) \end{bmatrix}$$

where $\mu_0 = -.72$, $\beta_0 = 3$, and $X_t$ and $Y_t$ are jointly i.i.d. random variables with distribution $\mathcal{N}(0, .16)$. The parameters are set to the usual values in the literature, see e.g., [35], [31], [73] and [56]. The [35] model is known to be challenging to estimate because it induces some instability for usual moment-based estimators in small samples. This kind of instability has been observed in several empirical applications.

Table 1 reports the mean-squarred error (MSE), bias and variance of the ESP and ET estimators for different sample sizes. The MSE of the ESP estimator is always smaller than for the ET estimator, and the differences are notable for small sample sizes. The decomposition of the MSE as the sum of the variance and the squared bias indicates that the variance contributes more than the bias to the reduction of the MSE for the ESP estimator. Note also that Table 1 understates the improvement delivered by the variance penalization of the ESP objective function. We helped the ET estimator (or equivalently, the MM estimator), by restricting its parameter space to $\beta < 15$. Without this parameter restriction, the behaviour of the ET estimator is very unstable for sample sizes below 100. An analysis of the typical shape of the objective functions for small sample

TABLE 1
***ESP vs. ET estimator for the two-parameter Hall and Horowitz model.***

| $T$ | | $\beta$ | | $\mu$ | |
|---|---|---|---|---|---|
| | | ET | ESP | ET | ESP |
| 25 | MSE | 3.6729 | 0.6644 | 1.5633 | 0.2356 |
| | Bias | 0.4761 | −0.0127 | −0.1955 | 0.1042 |
| | Var. | 3.4462 | 0.6642 | 1.5250 | 0.2247 |
| 50 | MSE | 1.5685 | 0.3359 | 0.8974 | 0.1234 |
| | Bias | 0.2602 | −0.0128 | −0.1232 | 0.0638 |
| | Var. | 1.5008 | 0.3358 | 0.8822 | 0.1193 |
| 100 | MSE | 0.6653 | 0.1589 | 0.4375 | 0.0611 |
| | Bias | 0.1490 | −0.0085 | −0.0770 | 0.0355 |
| | Var. | 0.6431 | 0.1589 | 0.4316 | 0.0599 |
| 200 | MSE | 0.2361 | 0.0823 | 0.1744 | 0.0312 |
| | Bias | 0.0633 | −0.0158 | −0.0314 | 0.0250 |
| | Var. | 0.2321 | 0.0821 | 0.1734 | 0.0305 |
| 500 | MSE | 0.0435 | 0.0322 | 0.0198 | 0.0134 |
| | Bias | 0.0229 | −0.0096 | −0.0102 | 0.0115 |
| | Var. | 0.0430 | 0.0322 | 0.0197 | 0.0133 |
| 1000 | MSE | 0.0243 | 0.0184 | 0.0115 | 0.0071 |
| | Bias | 0.0137 | −0.0038 | −0.0060 | 0.0060 |
| | Var. | 0.0241 | 0.0184 | 0.0115 | 0.0071 |
| 5000 | MSE | 0.0040 | 0.0038 | 0.0016 | 0.0015 |
| | Bias | 0.0016 | −0.0021 | −0.0005 | 0.0021 |
| | Var. | 0.0040 | 0.0038 | 0.0016 | 0.0015 |

Note: The reported statistics are based on 10,000 simulated samples of sample size equal to the indicated $T$. For ET, the parameter space is restricted to $\beta < 15$ in order to limit the erratic behaviour of the estimator at sample sizes $T = 25$ and 50. No such parameter restriction is imposed for ESP.

size explains this phenomenon. The typical ET objective function has a ridge that follows from around the population parameter values ($\beta_0 = 3$, $\mu_0 = -.72$) towards $(1000, -600)$. The ridgeline is not totally flat, and it often has a gentle downward slope as we move away from the area near the population parameter values. However, regularly, for some simulated samples, the very top of the ridge is extremely far from the population parameter values, so that ET estimates are very far from the population parameter values. This does not happen for the ESP estimator. The variance term of the ESP objective function ensures that the ridge drops sufficiently as we move away from the maximum that is near the population parameter value. Thus, in line with our finite-sample analysis of the ESP objective function (Section 2.2), the ESP estimator is much more stable.

#### *4.1.2. ET and ESP estimators for a stochastic volatility model*

The stochastic lognormal volatility model has been a competitor for the GARCH model. The system evolves as

$$\ln\left(\sigma_t^2\right) = w + \beta \ln\left(\sigma_t^2\right) + \sigma_u U_t$$

and

$$Y_t = \sigma_t Z_t$$

where $U_t$ and $Z_t$ are jointly i.i.d. random variables with distribution $\mathcal{N}(0,1)$. The model has been used to compare moment-based estimators [e.g., 3, 56].

TABLE 2
***ESP vs. ET estimator for the two-parameter stochastic volatility model.***

| $T$ | | $w$ | | $\sigma_u$ | |
|---|---|---|---|---|---|
| | | ET | ESP | ET | ESP |
| 25 | MSE | 0.0015 | 0.0016 | 0.0615 | 0.0187 |
| | Bias | −0.0048 | −0.0069 | −0.1549 | −0.0798 |
| | Var. | 0.0015 | 0.0015 | 0.0375 | 0.0123 |
| 50 | MSE | 0.0010 | 0.0010 | 0.0420 | 0.0158 |
| | Bias | −0.0009 | −0.0022 | −0.1214 | −0.0709 |
| | Var. | 0.0010 | 0.0010 | 0.0272 | 0.0108 |
| 100 | MSE | 0.0006 | 0.0006 | 0.0226 | 0.0113 |
| | Bias | 0.0001 | −0.0005 | −0.0814 | −0.0558 |
| | Var. | 0.0006 | 0.0006 | 0.0160 | 0.0082 |
| 200 | MSE | 3e-04 | 3e-04 | 1e-02 | 7e-03 |
| | Bias | 0.0005 | 0.0003 | −0.0471 | −0.0384 |
| | Var. | 0.0003 | 0.0003 | 0.0078 | 0.0055 |

Note: The reported statistics are based on 10,000 simulated samples of sample size equal to the indicated $T$.

As documented in [3], the joint estimation of $\beta$, $w$ and $\sigma_u$ yield numerical convergence problem. Thus, we fix $\beta = .95$ and estimate the two parameters $(w, \sigma_u)$ with the moment function

$$\psi_t(w,\sigma_u) = \begin{bmatrix} |Y_t| - \sqrt{\frac{2}{\pi}} \exp\left\{\frac{w}{2(1-.95)} + \frac{\sigma_u^2}{8(1-.95^2)}\right\} \\ Y_t^2 - \exp\left\{\frac{w}{(1-.95)} + \frac{\sigma_u^2}{2(1-.95^2)}\right\} \end{bmatrix}.$$

We simulate the model for $(w_0, \sigma_{u,0}) = (-0.368, 0.260)$ in order to match the middle case considered in [3].

Table 2 shows that the ESP MSE are either similar to, or smaller than, the ET MSE. For the $w$ parameter, beyond the similarity in terms of MSE, the biases are slightly different : The ESP bias is slightly bigger. For the $\sigma_u$ parameter, the ESP MSE, bias and variance are always smaller, although the difference is never big. Overall, the results indicate that the ET and the ESP estimators perform similarly, when the first one already performs well.

### *4.1.3. Test statistics for the two-parameter Hall and Horowitz model*

In the present section we investigate the finite-sample behavior of the trinity+1. We simulate again the two-parameter Hall and Horowitz model, and study the actual rejection probabilities of the test statistics for the null hypothesis $H_0 : \beta = 3$ and $\mu = -.72$, i.e., one minus the actual coverage probability. We do not report the actual rejection probability for the LM test, for which the asymptotic results of Theorem 2 provide poor finite sample guidance. As previously mentioned, LM tests are typically unreliable in nonlinear setting because of local extrema. For comparison, we also report the actual rejection probabilities for the ET ALR test statistic.

Table 3 presents the results. The performances of the different test statistics are comparable in terms of actual rejection probabilities, although the ESP $\mathrm{ALR}_T$ and the ESP $\mathrm{Tilt}_T$ seem to perform slightly better. The closer is the actual rejection probabilities to the nominal size $\alpha = .05$ the more accurate is the asymptotic approximation provided by Theorem 2.

TABLE 3
***Actual rejection probabilities for the two-parameter Hall and Horowitz model.***

| $T$ | ESP $\mathrm{ALR}_T$ | ET $\mathrm{ALR}_T$ | ESP $\mathrm{Wald}_T$ | ESP $\mathrm{Tilt}_T$ |
|---|---|---|---|---|
| 50 | 0.1996 | 0.1964 | 0.2332 | 0.2016 |
| 100 | 0.1608 | 0.1622 | 0.1831 | 0.1722 |
| 200 | 0.1265 | 0.1278 | 0.1419 | 0.1361 |
| 1000 | 0.0765 | 0.0783 | 0.0839 | 0.0821 |
| 2000 | 0.0669 | 0.0681 | 0.0688 | 0.0676 |

Note: Under the null hypothesis $H_0 : \theta = 3$ and $\mu = -.72$, asymptotically the test statistics follows a chi-square distribution with two degree of freedom. The tests used the critical value with size of $\alpha = .05$. The probabilities are based on 10,000 simulated samples of sample size equal to the indicated $T$.

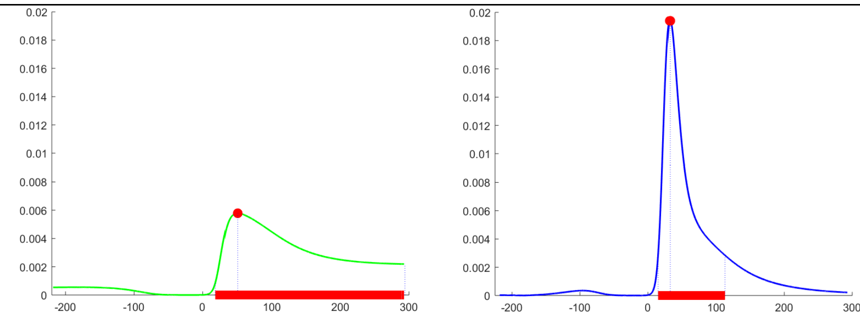
### *4.2. Empirical example*

In this section, we present an empirical example from asset pricing. In empirical consumption-based asset pricing, the literature has found little common ground about the value of the RRA of the representative agent: In most studies, point estimates from economically similar moment conditions are generally outside of each other's confidence intervals. The present section revisits the estimation of the RRA, which goes back to [38]. The popularity of moment-based estimation in consumption-based asset pricing, and more generally in economics is due to the fact that moment-based estimation does not necessarily require the specification of a family of distributions for the data [e.g. 36, sec. 3]. Typically, an economic model does not imply such family of distributions, except for tractability reasons. Imposing a family of distributions makes it difficult to disentangle the part of the inference results due to the empirical relevance of the economic model from the part due to these additional restrictions. Under regularity conditions, assuming a distribution corresponds to imposing an infinite number of extra moment restrictions [e.g., 25, chap. VII, sec. 3].

In order to estimate the RRA $\theta$, we rely on the following moment condition

$$\mathbb{E}\left[\left(\frac{C_t}{C_{t-1}}\right)^{-\theta}(R_{m,t} - R_{f,t})\right] = 0, \tag{10}$$

where $\frac{C_t}{C_{t-1}}$ is the growth consumption and $(R_{m,t} - R_{f,t})$ the market return in excess of the risk-free rate. The moment condition condition (10) and the data, which correspond to standard US data at yearly frequency from Shiller's website



(A) ET est. and ALR conf. region. (B) ESP est. and ALR conf. region.

Fig 1. ***ET vs. ESP inference (1890–2009)***

spanning from 1890 to 2009, are similar to [49]. The moment condition (10) has several advantages. Firstly, it is as consistent with [58] as with more recent consumption-based asset-pricing models, such as [4] or [29]. In other words, despite its simplicity it also correspond to sophisticated models, and it allows us to obtain estimates that are robust to different variations of consumption-based asset pricing theory. Secondly, without loss of generality, it does not require to estimate the time discount rate, about which there is little debate: The time discount rate of the representative agent is consistently found to be between .9 and 1.

In some of the more recent literature, it has been common to use other moment conditions with a separate parameter for the so-called intertemporal elasticity of substitution, i.e., use Epstein-Zin-Weil preferences [e.g. 20]. However, [5, 6] show that such a specification makes the economic interpretation of the parameters difficult. In particular, they show that an increase of the so-called RRA parameter does not yield a behaviour that would be considered more risk averse [5] E.g., All other things being equal, savings can be a decreasing function of the so-called RRA parameter for an agent with Epstein-Zin-Weil preferences [e.g., 6, sec. 6]. This difficulty of interpretation comes from a violation of the monotonicity axiom according to which an agent does not choose an action if another available action is preferable in every state of the world.

In light of Section 2, we report ET and ESP estimates as well as confidence regions based on the inversion of the ALR test statistics of Theorem 2 (p. 3680) with $\check{\theta}_T = \theta_0$. The latter have the advantage to better take into account the whole shape of the objective function than the other confidence regions such as the Wald-based (i.e., $t$-statistics-based) confidence regions, which only account for the shape of the objective function in a neighborhood of the estimate through its standard errors.

In Figure 1, (A) and (B) respectively displays the ET term and the ESP approximation. For ease of comparison, the scale is the same, and we normalize both of them so they integrate to one. The normalized ET term is much flatter around its maximum than the normalized ESP approximation. Flatness of the objective function around the estimate has been documented for other existing moment-based estimators, and it has often been regarded as one of the main sources of the instability of the RRA estimates [e.g., 34, p. 60-64]. Figure (B) shows that the normalized ESP is sharp around the ESP estimator. The relative sharpness of the ESP yields sharper confidence regions: The ESP confidence region is less than half its ET counterpart. In light of the variance penalization term in the ESP objective function (Section 2.2 on p. 3677) and the shrinkage-like behavior of the ESP estimator in the Monte-Carlo simulations (Section 4.1), the relative sharpness of the ESP inference is not surprising. In the Appendix section 11, additional empirical evidences corroborate the increased stability and precision of the ESP estimator w.r.t. the ET estimator (or equivalently, MM estimator).

## 5. Connection to the literature and further research directions

The present paper demonstrates a previously unknown connection between the SP approximation and moment-based estimation, and hence it is related to many papers in these literatures on top of the ones already cited. Following [15], the literature in statistics [e.g., 18, 79, 46, 83, 7, 24] and econometrics [e.g., 65, 42, 66, 55, 1] has used the SP (saddlepoint) and ESP approximations to obtain accurate approximations of distributions, especially in the tails. The strand of the SP literature that is closest to our paper derives SP approximations to the distribution of statistics that correspond to solutions of nonlinear estimating equations. The latter strand of literature started with [27] and continued with [76, 60, 43, 48, 2, 70], and [71], among others. More recently, [14], [59], and [56, 53, 54] propose more accurate tests for indirect inference, functional measurement error models, moment condition models, nonlinear estimators and GEL (generalized empirical likelihood) estimators, respectively. To the best of our knowledge, unlike the present paper, none of the prior papers use the SP or the ESP to develop an estimation method that yields a novel moment-based estimator. In ongoing work, we generalize the ESP approximation to the over-identified case and time-dependent data.

The present paper is also related to a large and growing literature on shrinkage estimators. Following [80]'s example, shrinkage has emerged as a powerful idea to develop more stable estimation methods. Examples include the ridge regression [39], the LASSO regression [82], the SCAD penalization [22, 23], and the elastic net penalization [87]. Some of the latter have been adapted and extended to moment-based estimation [e.g., 9]. While the ESP estimator can be regarded as a shrinkage estimator (Section 2.2), it has several particularities. Firstly, unlike the aforementioned shrinkage estimators, the ESP estimator does not require the calibration of tuning parameters, which is often delicate [e.g., 75]. Secondly, the ESP estimator does not require the user to choose a parameter value. The ESP estimator is *not* shrunk toward a user-chosen shrinkage value, but toward parameter values with lower estimated implied variance. Such a data-driven determination of the shrinkage value is particularly convenient for nonlinear moment-based estimation: While in regression models the choice of a shrinkage value is often easy to justify —e.g., zero, which corresponds to a more parsimonious model—, the choice is often more difficult for nonlinear moment-based estimation. E.g., in the numerical and the empirical example, it is unclear why one would like to shrink the risk-aversion parameter toward zero. Thirdly, the shrinkage nature of the ESP estimator is a consequence of defining an estimator that maximizes the ESP approximation of the finite-sample distribution of solutions to empirical moment conditions. It is not the consequence of the addition of an ad hoc penalization as is sometimes the case for shrinkage estimators.

As hinted in Remark 1 (p. 3675), the present paper is additionally related to Bayesian inference. Like several widely-used shrinkage estimators (e.g., 39, sec. 6; 82, sec. 5), the ESP estimator has connections to Bayesian inference. For example, the variance term of the ESP approximation share some similarities with the Jeffreys' prior used in parametric Bayesian inference. The investigation of these connections are left for future research. In a companion paper, we investigate the asymptotic connection between the ESP approximation and Bayesian posterior distributions.

Finally, the present paper is related to the econometric weak instrument literature, which is also motivated by the poor finite-sample stability and performance of usual moment-based estimators [e.g., see Introduction in 81, which is the seminal paper of the literature]. Despite a common motivation, there are major differences with respect to the present paper. Firstly, by definition, the weak instrument approach requires to assume that the moment conditions depend on the sample size $T$ [81, Assumption C]. In several applications (e.g., the empirical example in Section 4.2), this definitional assumption is incompatible with the model of interest. As [34] explains in his standard textbook on moment-based estimation, this definitional assumption is "artificial" in the sense that nobody seems to believe that economic and financial data induce moment conditions depending on the sample size $T$ in this way: This is just a "mathematical device" that is used to derive an asymptotic theory that aims at providing good approximations to finite sample behaviours. See also [81] for a similar justification of the definitional assumption. In contrast, no modification of the moment conditions is required for the ESP estimator. The idea is simply to define an estimator that maximizes an accurate approximation of the finite-sample distribution of the solutions to the empirical moment conditions. The relative stability and sharpness of the ESP objective function in both the numerical and the empirical examples illustrate the usefulness of the idea. Secondly, unlike the present paper, the weak instrument literature does not provide new estimators, but only novel test statistics. Actually, moment-based estimators are generally inconsistent for weakly identified parameters (81, Theorem 1; 32, Theorem 2). Thirdly, test statistics derived under the weak instrument assumption induce confidence regions that can be empty [e.g., 81, Table IV-VI], and that have infinite length with positive probability [17], so they can be "unreasonable" [61]. In contrast to Anderson-Rubin-type statistics used in the weak instrument literature [81], the test statistics of the present paper enjoy the same properties as the traditional trinity statistics, and thus do not yield empty confidence regions. Finally, note that, when, in applications, general test statistics robust to weak instrument appear necessary, it should be possible to extend the present paper for this purpose.

**Acknowledgments**

Parts of the present paper have previously circulated under the title "The Empirical Saddlepoint Likelihood Estimator Applied to Two-Step GMM" [78]. Some proofs of the present paper also borrow technical results from [40]. Helpful comments were provided by Patrick Gagliardini (discussant), Philipp Ketz (discussant), Eric Renault, Enrique Sentana, Aman Ullah and seminar/conference participants at Carnegie Mellon University, CFE-CMStatistics 2017, Swiss Finance Institute (EPFL and the University of Lausanne), 10th French Econometrics Conference (Paris School of Economics), at the Econometric Society European Winter Meeting 2018 (University of Naples Federico II), the University of Luxembourg, and Collegio Carlo Alberto (University of Torino, 2nd LTI conference).

**Supplementary Material**

**Appendix: The Empirical Saddlepoint Estimator**
(doi: 10.1214/21-EJS1976SUPP; .pdf). The Appendix mainly consists of a detailed proof of the first part of Theorem 1(i), i.e., existence and consistency of the ESP estimator. The proof relies on set-valued analysis. The high level of details should make the proof more transparent, ease the use of the intermediary results in further research, and make clear that the assumptions and the proofs available in the current literature are mathematically insufficient. The proofs of the other results (i.e., asymptotic normality of the ESP estimator, and Theorem 2, asymptotic distributions of the Trintiy+1 test statistics) are skipped because, while technical, they rely on extensions of more standard arguments, so the indications in the main text should be sufficient. Nevertheless, the latter proofs are available in [41]. The appendix contains a formalization of the variance shrinkage, an additional numerical example, and complementary information regarding the numerical and empirical examples.

**References**

[1] Aït-Sahalia, Y. and Yu, J. (2006). Saddlepoint approximations for continuous-time Markov processes. *Journal of Econometrics* **134** 507–551. MR2327401

[2] Almudevar, A., Field, C. and Robinson, J. (2000). The Density of Multivariate $M$-Estimates. *The Annals of Statistics* **28** 275–297. MR1762912

[3] Andersen, T. G. and Sørensen, B. E. (1996). GMM Estimation of a Stochastic Volatility Model: A Monte-Carlo Study. *Journal of Business & Economic Statistics* **14** 328–352.

[4] Barro, R. J. (2006). Rare Disasters and Asset Markets in the Twentieth Century. *The Quarterly Journal of Economics* **121** 823–866.
[5] Bommier, A., Chassagnon, A. and Legrand, F. (2012). Comparative Risk Aversion : A Formal Approach with Applications to saving behavior. *Journal of Economic Theory* **147** 1614–1641. MR2942625
[6] Bommier, A., Kochov, A. and Legrand, F. (2017). On Monotone Recursive Preferences. *Econometrica* **85** 1433–1466. MR3708628
[7] Broda, S. and Kan, R. (2015). On Distributions of Ratios. *Biometrika* **103** 205–218. MR3465831
[8] Butler, R. W. (2007). *Saddlepoint Approximations with Applications*. *Series in Statistical and Probabilistic Mathematics*. Cambridge University Press. MR2357347
[9] Caner, M. (2009). Lasso-Type GMM Estimator. *Econometric Theory* **25** 270–290. MR2472053
[10] Carrasco, M. and Florens, J.-P. (2014). On the asymptotic efficiency of GMM. *Econometric Theory* **30** 372–406. MR3231496
[11] Chamberlain, G. (1987). Asymptotic efficiency in estimation with conditional moment restrictions. *Journal of Econometrics* **34** 305–334. MR0888070
[12] Chib, S., Shin, M. and Simoni, A. (2018). Bayesian Estimation and Comparison of Moment Condition Models. *Journal of the American Statistical Association* **113** 1656–1668. MR3902236
[13] Csiszár, I. (1975). I-divergence geometry of probability distributions and minimization problems. *The Annals of Probability* 146–158. MR0365798
[14] Czellar, V. and Ronchetti, E. (2010). Accurate and robust tests for indirect inference. *Biometrika* **97** 621–630. MR2672487
[15] Daniels, H. E. (1954). Saddlepoint Approximations in Statistics. *The Annals of Mathematical Statistics* **25** 631–650. MR0066602
[16] Davison, A. C. and Hinkley, D. V. (1988). Saddlepoint Approximations in Resampling Methods. *Biometrika* **75** 417–431. MR0967581
[17] Dufour, J.-M. (1997). Some Impossibility Theorems in Econometrics With Applications to Structural and Dynamic Models. *Econometrica* **65** 1365–1387. MR1604305
[18] Easton, G. S. and Ronchetti, E. (1986). General Saddlepoint Approximations with Applications to L Statistics. *Journal of the American Statistical Association* **81** 420–430. MR0845879
[19] Efron, B. (1981). Nonparametric Standard Errors and Confidence Intervals. *The Canadian Journal of Statistics / La Revue Canadienne de Statistique* **9** 139–158. MR0640014
[20] Epstein, L. G. and Zin, S. E. (1991). Risk Aversion, and the Temporal Behavior of Consumption and Asset Returns: An Empirical Analysis. *The Journal of Political Economy* **99** 263–286. MR1006550
[21] Esscher, F. (1932). On the probability function in the collective theory of risk. *Scandinavian Actuarial Journal* 175–195.
[22] Fan, J. (1997). COMMENTS ON "WAVELETS IN STATISTICS :A REVIEW" BY A. ANTONIADIS. *Journal of the Italian Statistical Associa-*

*tion* **2** 131–138. MR1946364

[23] Fan, J. and Li, R. (2001). Variable Selection via Nonconcave Penalized Likelihood and Its Oracle Properties. *Journal of the American Statistical Association* **96** 1348–1360. MR1946581

[24] Fasiolo, M., Wood, S. N., Hartig, F. and Bravington, M. V. (2018). An extended empirical saddlepoint approximation for intractable likelihoods. *Electronic Journal of Statistics* **12** 1544–1578. MR3806432

[25] Feller, W. (1966/1971). *An Introduction to Probability Theory and Its Applications.* **2**, 2nd ed. Wiley. MR1534302

[26] Feuerverger, A. (1989). On The Empirical Saddlepoint Approximation. *Biometrika* **76** 457–464. MR1040640

[27] Field, C. (1982). Small Sample Asymptotic Expansions for Multivariate M-Estimates. *The Annals of Statistics,* **10** 672–689. MR0663425

[28] Field, C. A. and Ronchetti, E. (1990). *Small Sample Asymptotics. Lecture notes-Monograph Series*. Institute of Mathematical Statistics. MR1088480

[29] Gabaix, X. (2012). Variable Rare Disasters: An Exactly Solved Framework For Ten Puzzles in Macro-Finance. *The Quarterly Journal of Economics* **127** 645–700.

[30] Goutis, C. and Casella, G. (1999). Explaining the Saddlepoint Approximation. *The American Statistician* **53** 216–224. MR1711555

[31] Gregory, A. W., Lamarche, J.-F. and Smith, G. W. (2002). Information-theoretic estimation of preference parameter: macroeconomic applications and simulation evidence. *Journal of Econometrics* **107** 213-233. MR1889960

[32] Guggenberger, P. and Smith, R. J. (2005). Generalized Empirical Likelihood Estimators and Tests Under Partial, Weak, and Strong Identification. *Econometric Theory* **21** 667–709. MR2189492

[33] Haberman, S. J. (1984). Adjustmennt by Minimum Discriminant Information. *The Annals of Statistics* **12** 971–988. MR0751286

[34] Hall, A. R. (2005). *Generalized Method of Moments*. *Advanced Texts in Econometrics*. Oxford University Press. MR2135106

[35] Hall, P. and Horowitz, J. L. (1996). Bootstrap Critical Values for Tests Based on Generalized-Method-of-Moments Estimators. *Econometrica* **64** 891–916. MR1399222

[36] Hansen, L. P. (2013). Uncertainty Outside and Inside Economic Models In *Nobel Prize Lecture*.

[37] Hansen, L. P., Heaton, J. and Yaron, A. (1996). Finite-Sample Properties of Some Alternative GMM Estimators. *Journal of Business and Economic Statistics* **14** 262–280.

[38] Hansen, L. P. and Singleton, K. J. (1982). Generalized Instrumental Variables Estimation of Nonlinear Rational Expectations Models. *Econometrica* **50** 1269–1286. MR0673689

[39] Hoerland, A. E. and Kennard, R. W. (1970). Ridge Regression: Biased Estimation for Nonorthogonal Problems. *Technometrics* **12** 55–67.

[40] Holcblat, B. (2012). A Classical Moment-Based Inference Frame-

work with Bayesian Properties, PhD thesis, Carnegie Mellon University. MR3067857

[41] Holcblat, B. and Sowell, F. (2019). Complete Online Appendix to "The ESP estimator". Available online at https://arxiv.org/abs/1905.06977.

[42] Holly, A. and Phillips, P. C. B. (1979). A Saddlepoint Approximation to the Distribution of the $K$-Class Estimator of a Coefficient in a Simultaneous System. *Econometrica* **47** 1527–1547. MR0550946

[43] Imbens, G. W. (1997). One-Step Estimators for Over-Identified Generalized Method of Moments Models. *The Review of Economic Studies* **64** 359–383. MR1456135

[44] Imbens, G. W., Spady, R. H. and Johnson, P. (1998). Information Theoretic Approaches to Inference in Moment Condition Models. *Econometrica* **66** 333–357. MR1612246

[45] Jennrich, R. I. (1969). Asymptotic Properties of Non-Linear Least Squares Estimators. *The Annals of Mathematical Statistics* **40** 633–643. MR0238419

[46] Jensen, J. L. (1992). The Modified Signed Likelihood Statistic and Saddlepoint Approximations. *Biometrika* **79** 693–703. MR1209471

[47] Jensen, J. L. (1995). *Saddlepoint Approximations. Oxford Statistical Science Series*. Oxford University Press. MR1354837

[48] Jensen, J. L. and Wood, A. T. A. (1998). Large Deviation and Other Results For Minimum Contrast Estimators. *Annals of the Institute of Statistical Mathematics* **50** 673–695. MR1671986

[49] Julliard, C. and Ghosh, A. (2012). Can Rare Events Explain the Equity Premium Puzzle? *Review of Financial Studies* **25**.

[50] Kitamura, Y. (2001). Asymptotic Optimality of Empirical Likelihood for Testing Moment Restrictions. *Econometrica* **69** 1661–1672. MR1865225

[51] Kitamura, Y. and Stutzer, M. (1997). An Information-Theoretic Alternative to Generalized Method of Moments Estimation. *Econometrica* **65** 861–874. MR1458431

[52] Kolassa, J. E. (1994/2006). *Series Approximation Methods in Statistics. Lecture Notes in Statistics* **88**. Springer.

[53] Kundhi, G. and Rilstone, P. (2013). Edgeworth and Saddlepoint Expansions for Nonlinear Estimators. *Econometric Theory* **29** 1057–1078. MR3148824

[54] Kundhi, G. and Rilstone, P. (2015). Saddlepoint expansions for GEL estimators. *Statistical Methods & Applications* **24** 1–24. MR3328164

[55] Lieberman, O. (1994). On the Approximation of Saddlepoint Expansions in Statistics. *Econometric Theory* **10** 900–916. MR1323281

[56] Lô, S. N. and Ronchetti, E. (2012). Robust small sample accurate inference in moment condition models. *Computational Statistics and Data Analysis* **56** 3182–3197. MR2943888

[57] Loader, C. (2000). Fast and Accurate Computation of Binomial Probabilities.

[58] Lucas, R. E. (1978). Asset Prices in an Exchange Economy. *Econometrica*

**46** 1429–1445. MR0513698

[59] Ma, Y. and Ronchetti, E. (2011). Saddlepoint Test in Measurement Error Models. *Journal of the American Statistical Association* **106** 147–156. MR2816709

[60] Monti, A. C. and Ronchetti, E. (1993). On the relationship between empirical likelihood and empirical saddlepoint approximation for multivariate M-estimators. *Biometrika* **80**. MR1243508

[61] Müller, U. K. and Norets, A. (2016). Credibility of confidence sets in nonstandard econometric problems. *Econometrica* **84** 2183–2213. MR3580266

[62] Newey, W. K. and Smith, R. J. (2004). Higher Order Properties of GMM and Generalized Empirical Likelihood Estimators. *Econometrica* **72** 219–255. MR2031017

[63] Pearson, K. (1894). Contribution to the mathematical theory of evolution. *Philosophical Transactions of the Royal Society* 71–110.

[64] Pearson, K. (1902). On the systematic fitting of curves to observations and measurements, Parts I and II. *Biometrika* **1, 2** 265–303, 1–23.

[65] Phillips, P. C. B. (1978). Edgeworth and saddlepoint approximations in the first-order non circular autoregression. *Biometrika* **65** 91–98.

[66] Phillips, P. C. B. (1982). Exact Small Theory in the Simultaneous Equations Model. *Cowles Foundation Discussion Paper NO. 621*.

[67] Qin, J. and Lawless, J. (1994). Empirical Likelihood and General Estimating Equations. *The Annals of Statistics* **22** 300–325. MR1272085

[68] Rilstone, P., Srivastava, V. K. and Ullah, A. (1996). The second-order bias and mean squared error of nonlinear estimators. *Journal of Econometrics* **75** 369–395. MR1413647

[69] Robert, C. P. (2007/1994). *The Bayesian Choice. From Decision-Theoretic Foundations to Computational Implementation*, second ed. *Texts in Statistics*. Springer. MR2723361

[70] Robinson, J., Ronchetti, E. and Young, G. A. (2003). Saddlepoint Approximations and Tests Based on Multivariate $M$-Estimates. *Annals of Statistics* **31** 1154–1169. MR2001646

[71] Ronchetti, E. and Trojani, F. (2003). Saddlepoint Approximations and Test Statistics for Accurate Inference in Overidentified Moment Conditions Models. *Working paper, National Centre of Competence in Research, Financial Valuation and Risk Management*.

[72] Ronchetti, E. and Welsh, A. H. (1994). Empirical Saddlepoint Approximations for Multivariate M-Estimators. *Journal of the Royal Statistical Society.Series B (Methodological),* **56** 313–326. MR1281936

[73] Schennach, S. (2007). Point Estimation with Exponentially Tilted Empirical Likelihood. *The Annals of Statistics* **35** 634–672. MR2336862

[74] Schmetterer, L. (1966). *Mathematische Statistik*, Second ed. Springer.

[75] Sengupta, N. and Sowell, F. (2020). On the Asymptotic Distribution of Ridge Regression Estimators Using Training and Test Samples. *Econometrics* **8**. doi:10.3390/econometrics8040039.

[76] Skovgaard, I. M. (1990). On the Density of Minimum Contrast Estima-

tors. *The Annals of Statistics* **18** 779–789. MR1056336

[77] Smith, R. J. (2011). GEL criteria for moment condition models. *Econometric Theory* **27** 1192–1235. MR2868838

[78] Sowell, F. (2009). The Empirical Saddlepoint Likelihood Estimator Applied to Two-Step GMM. *working paper, Tepper School of Business, Carnegie Mellon University.*

[79] Spady, R. H. (1991). Saddlepoint Approximations for Regression Models. *Biometrika* **78** 879–889. MR1147025

[80] Stein, C. (1956). Inadmissibility of the Usual Estimator for the Mean of a Multivariate Normal Distribution. In *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability* **Volume 1: Contributions to the Theory of Statistics** 197–206. University of California Press, Berkeley, California. MR0084922

[81] Stock, J. H. and Wright, J. H. (2000). GMM with Weak Identification. *Econometrica* **68** 1055–1096. MR1779144

[82] Tibshirani, R. (1996). Regression Shrinkage and Selection via the Lasso. *Journal of the Royal Statistical Society. Series B (Methodological)* **58** 267–288. MR1379242

[83] Vecchia, D. L., Ronchetti, E. and Trojani, F. (2012). Higher-Order Infinitesimal Robustness. *Journal of the American Statistical Association* **107** 1546–1557. MR3036415

[84] Wald, A. (1949). Note on the Consistency of the Maximum Likelihood Estimate. *The Annals of Mathematical Statistics* **20** 595–601. MR0032169

[85] Wang, S. (1990). Saddlepoint approximations in resampling analysis. *Annals of the Institute of Statistical Mathematics* **42** 115–131. MR1054727

[86] Young, G. A. and Daniels, H. E. (1990). Bootstrap Bias. *Biometrika* **77** 179–185. MR1049419

[87] Zou, H. and Hastie, T. (2005). Regularization and Variable Selection via the Elastic Net. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)* **67** 301–320. MR2137327